

## ÖĞRENCİ DEĞERLENDİRMELERİ VE ÖĞRETMENLERİN ÖZ DEĞERLENDİRMELERİ ARASINDAKİ İLİŞKİ

### THE RELATIONSHIP BETWEEN STUDENT EVALUATION OF LECTURER PERFORMANCE AND LECTURER SELF- ASSESSMENT

*Özgür ŞAHAN, Kari Elizabeth ŞAHAN*

**ÖZET:** Öğrenciler, öğretmenlerin mesleki performanslarını bütün yönleriyle değerlendiremeyeceği için öğrenci değerlendirmeleri geçerlilik ve güvenilirlik açısından tartışmalı bir konudur. Buna rağmen öğrenci değerlendirmeleri tüm dünyada öğretmenlerin mesleki performanslarının değerlendirilmesinde kullanılan en yaygın ölçme araçlarından biridir. Öte yandan öğretmenlerin öz değerlendirmeleri ve öğrencilerin öğretmenlere vermiş oldukları puanların korelasyonu araştırmaya değer bir konudur. Bu çalışmada Bursa Teknik Üniversitesi Yabancı Diller Yüksekokulu tarafından geliştirilmiş performans değerlendirme anketi kullanılarak öğrencilerin öğretmenleri ve öğretmenlerin kendi öz değerlendirmeleri arasındaki ilişkiyi araştırmak hedeflenmektedir. Bu çalışmaya 139 öğrenci ve 19 öğretmen katılmıştır. Aynı zamanda bu çalışma öğretmenlerin öğrenci değerlendirmeleri açısından kendi yeterliliklerinin ne kadar farkında olduklarını da ortaya çıkarmayı amaçlamaktadır. Elde edilen sonuçlar öğretmenlerin ulusları ve tecrübeleri göz önünde bulundurularak değerlendirilmiştir. Çalışmanın ana bulgusu yansıtıcı öğretmenin önemini ortaya koymakla birlikte öğretmenlerin mesleki gelişimleri açısından mesleki yeterlilikleri ve performanslarının farkında olmalarının gerekliliğini vurgulamaktadır.

**Anahtar sözcükler:** Öğrenci değerlendirmesi, Öğretmen öz değerlendirmesi, Yansıtıcı öğretme

**ABSTRACT:** Student evaluations are the most commonly used tool for assessing lecturer performance and competencies throughout the world. Student evaluations continue to be popular despite objections in terms of reliability and validity check issues, since students may not be able to assess lecturer effectiveness in every aspect. Regarding lecturer self-assessment, it would be intriguing to investigate the correlations between lecturers' and students' perceptions of lecturer performance. This study aims to find out what students think about their lecturers' performance along with lecturers' self-perceptions of themselves using the same poll, specifically the poll developed by Bursa Technical University School of Foreign Languages. The participants consist of 139 students and 19 lecturers. The present study also intends to see how honest the lecturers are with themselves in terms of teaching competencies and performance with respect to student evaluations. The results are analyzed and discussed regarding lecturers' nationality (Turkish and foreign) and teaching experience. The main conclusion of the study emphasizes the importance of reflective-teaching and underlines how significant it is for lecturers to be aware of their performances and capabilities for self-improvement.

**Keywords:** Student evaluation, Lecturer self-assessment, Reflective teaching

#### 1. INTRODUCTION

Student evaluations of teaching (SET) may be the most important and commonly used tool to measure teachers' performances in most educational institutions throughout the world; they are seen as one of the best tools in terms of being supported by studies conducted in the relevant literature (Marsh, 1987; McKeachie, 1997; Seldin, 1999). Even with this acknowledgement, there exist many controversial issues about student evaluations. While there is a massive body of literature on student evaluations of teacher performance, most studies focus

on the validity of student ratings, arguing that students are unable to accurately assess teacher performance (McKeachie; Theall & Franklin, 2001). The inability of students to effectively measure teacher performance essentially turns student evaluations into a popularity contest (Beran & Rokosh, 2009). Additionally some studies are concerned with the reliability of this kind of instrument, which often has been found to lack adequate assistance in terms of practical teaching improvement (Apollonia & Abrami, 1997; Beran & Rokosh; Marsh, 1984). Another research topic concerning student evaluations is student response rate (Dommeyer, Baum, Hanna & Chapman, 2004; Nulty, 2007). Researchers have found that students display signs of fatigue or disinterest, and they are less likely to participate in teacher evaluations as they progress in their academic studies (Chen & Hoshower, 2003). Moreover, it has been found that response rates are lower for online evaluations than classic pen-and-paper in-class evaluations (Avery, Bryant, Kang & Bell, 2006).

Another issue that needs to be taken into consideration is the purpose that SETs serve. Berk (2005) defines two types of purposes—formative and summative—for measuring teaching effectiveness. The former gathers data to improve the quality of teaching, while the latter uses data to assess overall performance. Because they can be used for administrative purposes, such as tenure, merit, and promotion decisions (Kelly, 2012), student evaluations are more commonly used for summative rather than formative purposes (Gravestock & Gregor-Greenleaf, 2008). However, many researchers qualify that student ratings alone should not be used for summative decisions (Arreola, 2000; Braskamp & Ory, 1994; Seldin, 1999). In terms of the formative purposes of SETs, Kelly also draws attention to the timing of SETs which are generally conducted at the end of the semester, thereby giving no opportunities for teachers to reflect on their actions and make changes in their teaching methods, thereby limiting the useful potential of student evaluations.

In the spirit of reviewing multiple tools for accurate teaching evaluation, this study aims to examine the relationship between student evaluations of teacher performance and teacher self-assessments of their teaching performance. Therefore, the current study asks the following research questions:

1. Is there any correlation between student evaluations and teacher self-assessment through the same performance survey?
2. Do students think teachers who share the same L1 background (Turkish) as them are better teachers than foreign EFL teachers, including native speakers?
3. Do the self-evaluations of Turkish and foreign EFL teachers differ in terms of self-perceived teaching effectiveness?

### **1.1. Literature Review**

There is no doubt that students score effective teachers higher than less effective teachers (Bélanger & Longden, 2009). However, effective teaching is a sophisticated issue; therefore, a variety of factors influence student evaluations of teaching. Some studies (Griffin, 2004; Marsh & Roche, 2000; Olivares, 2001) have underlined a positive correlation between student academic success and SET scores. In another study, Naftulin, Ware, and Donnelly (1973) highlight the influence that a teachers' sense of humor has on student attitudes towards that teacher, concluding that teachers who entertain students get higher SET scores. Moreover, some external factors, such as the course's status—either elective or required—may have an effect on SET scores (McKeachie, 1979), and class size may impact the ratings (Bedard & Kuhn, 2008), playing a prominent role on the overall evaluations.

To counteract the biasing effects of external factors, the evidence of teaching effectiveness can be drawn from different sources with different types of measurements. Berk (2005, p. 49)

### The Relationship between Student Evaluation of Lecturer Performance and Lecturer Self-Assessment

lists these sources as: 1) student ratings, 2) peer ratings, 3) self-evaluation, 4) videos, 5) student interviews, 6) alumni ratings, 7) employer ratings, 8) administrator ratings, 9) teaching scholarship, 10) teaching awards, 11) learning outcome measures, and 12) teaching portfolio. He recommends the triangulation of these sources to produce more meaningful evidence to account for the complex nature of teaching. He also suggests the use of multiple sources, specifically noting that different sources build upon the strengths and compensate for the weaknesses of each other.

Martinson and Ryan (1981) conducted a research study on teacher performance evaluation with 144 Journalism students, half of whom completed the survey. The findings of the study showed that students were careful at being fair and serious in evaluations and avoided favoring popular, entertaining or easy grading teachers in their evaluations. Similar finding were corroborated in a more recent study conducted by Spencer and Schmelkin (2002). However, the students in the Martinson and Ryan study voiced their concerns about the instrument's validity and reliability along with the notion that student responses were disregarded by teachers, prohibiting them from making required changes. In the same vein, in a study conducted by Chen and Hoshower (1998), in which a total number of 208 students participated, expectancy theory was used to shift the researcher's focus to two important points: how students expected their feedback to be used and whether they expected it to be put into practice. The study found that the students' primary motivation for participating in the evaluations was to improve teaching quality in the classrooms, indicating that students are concerned more with formative than summative purposes for teacher assessment.

Considering teacher self-examinations, Wheeler and Knoop (1982) found that, in terms of performance assessment, teachers tended to give themselves higher ratings than their supervisors. In their study, which surveyed 47 student-teachers and their academic and field supervisors, the researchers found that while teachers tended to give themselves higher overall ratings, they also tended to display greater differentiation between items on the survey than their supervisors did. Rather, both their academic and field supervisors tended to evaluate the teachers on overall impression, providing little information about individual categories for teacher improvement and creating a halo effect. The study suggests that although teachers tend to be more lenient in their self-assessment, their ratings also tend to provide the greatest potential for improving their own teacher effectiveness.

In Beran and Rokosh's study (2009), the perceptions of 357 instructors in a Canadian university about the strengths and weaknesses of a student rating instrument were examined. Most of the participants in the study had negative views about the rating instrument, administration procedure, and use of SET results. They also thought that ratings may have negative effects on instructors and may build biases. Some instructors, however, supported the validity of the ratings, as 25% of the instructors considered ratings useful for the improvement of teaching effectiveness, causing the study to suggest that the survey measures should assess what instructors consider to be quality teaching.

Students' perspectives about student evaluations of teaching have also been in the center of research. In a recent study, Harris and Twinaime (2012) used a thematic analysis of students' reflective comments to see whether students value evaluation for teacher development. A total number of 59 students participated in the study at a university in New Zealand to assess the institution's evaluation system. According to the students in this study, students believe that evaluations are used for three main purposes: "1) development of the paper and the teacher, 2) the benefit of future students, and 3) fostering, increasing and reinforcing the communication between students and teachers regarding teaching and learning" (p. 155).

In a recent study, van Diggelen, den Brok and Beijard (2012) examined peer-assessment by looking at the process by which teachers gave feedback to and accepted it from their

colleagues. In the study, they evaluated teaching performance through assessment forms, video-taped feedback conversations and written reflective reports by 24 teachers. The findings show that teachers gave overall higher scores to themselves than their peers gave to them. Moreover, the study also found that teachers gave constructive and formative peer feedback that was generally accepted by other teachers. Additionally, the researchers found that the feedback was most productive when the teachers used clear criteria and standards to assess teaching performance.

All in all, SETs have been used as a prominent assessment tool to provide feedback for teachers. Although reliability and validity concerns exist about student evaluations, teachers may not be aware of their actual teaching performances and may reward themselves with higher or lower ratings in self-assessments than students would. Comparing SET scores and teacher self-assessment results may help teachers raise their own awareness of their capabilities and reflect on their actions for better teaching.

## **2. METHODOLOGY**

### **2.1. Participants**

A total number of 19 EFL lecturers (8 males and 11 females) participated in this study. Among the lecturers, 12 of them were Turkish and 7 of them were foreign teachers coming from different countries including America (n = 5), England (n= 1), and Kosovo (n = 1). The Turkish lecturers who participated in this experimental study were professionals in the field of interdisciplinary English language teaching, learning and assessment, and regular employees of the School of Foreign Languages at Bursa Technical University. They were all graduates from different English Language Teaching and English Language Literature departments in Turkey, and they had the same L1 background (Turkish). These 12 Turkish teachers all had various teaching backgrounds, and all at least held an MA degree in the field of interdisciplinary EFL and had at least three or more years of experience in EFL instruction. Two of the total 7 foreign lecturers were Fulbright scholars and the other foreign lecturers were hired with annual contracts by the institution. They had BA degrees in different fields other than language teaching, but all the foreign teachers were certified with TESOL or TEFL and had at least two or more years of EFL/ESL teaching experience.

The student participants in the study included a total number of 139 students, 16 of whom were pre-intermediate level students, 117 of whom were intermediate level students, and 6 of whom were upper intermediate level students. The students, who were all prep-class science-track students studying at Bursa Technical University's School of Foreign Languages in Turkey, were selected through the convenience sampling method. To be placed in this school, they took a public placement university examination and were ranked according to the scores from top to down.

The educational system of the School of Foreign Languages consists of 4 quarters based on 4 distinct language levels. All students attending the school (except for the students who earned a score of at least 65 percent on a nation-wide foreign language placement test, such as YDS) take a placement test at the beginning of the academic year, and they are put in classes according to the score they receive on the placement test, ranging from elementary to upper-intermediate. Each class in all levels includes at most 15 students and in each level students are educated for 30 hours weekly.

### **2.2. Data Collection Procedure**

The data for this study were collected during the first week of April 2014, which was the last class week of the third quarter. Students of BTU School of Foreign Languages are regularly requested to evaluate their lecturers' teaching performances, curricula and testing unit regularly

## **The Relationship between Student Evaluation of Lecturer Performance and Lecturer Self-Assessment**

at the end of each 8-week language level instruction period, called a quarter in the school system explained above. The results of the evaluations are reported to the lecturers before the next instruction period. Differently from this routine procedure, for the study, the lecturers were also requested to assess their performances on the same five-point Likert type scale. The only change was a shift in the subject of the statements from 'the lecturer' to the first-person pronoun, 'I'. The reports were presented to the lecturers in a comparative manner to underline the differences between their self-assessment ratings and the students' evaluations. The confidentiality of both the students' and the lecturers' ratings were assured to the participants. The data obtained from both students and lecturers constituted the basis of the present study.

### **2.3. Data Collection Instruments**

#### **2.3.1. The Lecturer Performance Scale**

The lecturer performance scale (see Appendix A) used in this study was developed by the Board of BTU School of Foreign Languages considering the dynamics and regulations of the institute and expectations from the teaching staff. The items were written on the poll in both Turkish and English to obtain more valid and reliable results from the student participants. The performance scale consisted of 20 items, the first 13 of which were directly about the lecturers' teaching performance inside and outside the classrooms, such as giving adequate answers to the questions asked by the students, using the teaching tools effectively, treating all students with respect and without any sort of offensive behavior, etc. The other 7 items were related to the curricula and testing issues. The participants were asked to indicate to what extent they agree on the items, and a lower rate reflected greater disagreement with the statement (e.g. "The lecturer/I provide(s) a suitable atmosphere for student participation in class", 1= strongly disagree). At the end of the scale a blank box was provided to the students to write comments. However, only the data gained from the first 13 items, which were directly related to the lecturers' professional performances, were used in the present study.

### **2.4. Data Analysis Procedure**

Since the present study intended to explore the influence of different independent variables (teaching experience and nationality) on one dependent variable (SET scores), two-way analysis of variance (ANOVA) was used to analyze the data gathered from the survey. After the analysis was completed, the correlation of the independent and dependent variables and the mean scores of the SET and self-assessments were obtained. The data were then organized and presented in the tables depicted in the following section.

## **3. RESULTS**

The data obtained from the students' evaluations of their lecturers' performances and the lecturers' self-assessments were analyzed through SPSS 20.0. In the study the lecturers' results are analyzed from two main perspectives: their teaching experience and their nationality. Teaching experience is defined as 'low' and 'high,' dividing the teachers into two categories. Adding the value given for each item of the survey, the highest total score is 65. The mean scores totaled from these values is the criterion of our comparisons in Table 1 and Table 2, which are shown below and from which the results for this survey are interpreted.

**Table 1: Mean Scores of Lecturers' Self-Assessment Ratings**

				$\bar{X}$
<b>Nationality</b>	<b>Experience</b>	<b>N</b>	<b>SD</b>	
Turkish	Low	6	5.456	55.17
	High	6	3.450	58.50
	Total	12	4.687	56.83
Foreign	Low	4	4.193	53.75
	High	3	1.155	54.67
	Total	7	3.078	54.14
Total	Low	10	4.789	54.60
	High	9	3.383	57.22
	Total	19	4.285	55.84

According to Table 1, there is no significant difference between Turkish and foreign lecturers in terms of their self-assessment mean scores. However, slight differences between mean scores may indicate that Turkish lecturers see themselves as more qualified ( $X = 56.83$ ) than foreign lecturers do ( $X = 54.14$ ). Besides the total mean scores, similar results can be observed in favor of Turkish lecturers in both the low and high experienced lecturer groups. Disregarding nationality, the results may indicate that high-experienced lecturers ( $X = 57.22$ ) see themselves as more capable and skillful than low-experienced lecturers do ( $X = 54.60$ ).

**Table 2: Mean Scores of Students' Evaluations of Lecturers' Performances**

<b>Nationality</b>	<b>Experience</b>	<b>N</b>	<b>SD</b>	<b><math>\bar{X}</math></b>
Turkish	Low	6	4.277	56.14
	High	6	5.600	58.30
	Total	12	4.883	57.22
Foreign	Low	4	10.002	52.98
	High	3	7.563	52.35
	Total	7	8.319	52.71
Total	Low	10	6.794	54.87
	High	9	6.539	56.31
	Total	19	6.529	55.56

When student ratings for lecturer performance were analyzed, no remarkable difference could be observed between Turkish and foreign lecturers in terms of the students' ratings for their lecturers' teaching performances (see Table 2). Considering the mean scores for each group of lecturers (low-experienced and high-experienced), the results may indicate that students perceive their Turkish lecturers ( $X = 57.22$ ) as more qualified and skillful than their foreign lecturers ( $X = 52.71$ ). Just as the high-experienced lecturers gave themselves higher points on the self-evaluation than low-experienced lecturers did, the students also see their high-experienced teachers ( $X = 56.31$ ) as more capable and skillful than their low-experienced teachers ( $X = 54.87$ ).

**Table 3: Correlations of Students' and Lecturers' Ratings (N = 19)**

		Nationality	Experience	Self-Assessment	Student Evaluation
Nationality	Correlation		-.069	-.311	-.342
	Sig.	--	.779	.195	.152
Experience	Correlation	-.069		.314	.113
	Sig.	.779	--	.191	.645
Self-Assessment	Correlation	-.311	.314		-.152
	Sig.	.195	.191	--	.535
Student Evaluation	Correlation	-.342	.113	-.152	
	Sig.	.152	.645	.535	--

Considering the degree of relationship between the variables under investigation, the correlation coefficient changes between -1 and +1, referring to the range of values which indicate the direction of correlation (negative or positive). As the strength of the relationship between the variables increases, so does the value of the correlation coefficient, with a value of 1 showing a perfect relationship. According to the correlation coefficients in Table 3, it is difficult to claim a significant correlation between variables. Each correlation among the variables under investigation was found to be approximately between -.400 and +.400 indicating a weak and meaningless relationship.

#### 4. DISCUSSION AND CONCLUSION

This study examined the relationship between teacher self-assessments and student evaluations of teaching performance in order to see how accurately lecturers were aware of their teaching performance with respect to student ratings. The results can be discussed in terms of three research questions of the study.

Considering the first research question, no correlation was observed between student ratings of their lecturers' performance and lecturer self-assessment scores given through the same survey instrument. However, when comparing the mean SET scores and self-assessment ratings, rather similar results were observed, which suggests that lecturers perceive their performances in a manner similar to student perceptions. This may imply that both students and teachers share similar criteria and expectations in terms of effective teaching performance. That is to say, lecturers may be relatively accurate in their self-evaluations because both lecturers and students hold common norms to define good teaching. In their study conducted in Finland, Kasanen and Rätty (2002) concluded that students learn to self-assess by replicating the criteria assumed by the school and conveyed through the teacher; students came to understand good classroom performance through the actions and reactions of their teacher. Similarly, although the present study looked at the self-evaluation scores of lecturers rather than students, it also found that lecturers and students share comparable assessment criteria demonstrating what the school values as important. Moreover, student evaluations may be a useful tool that allows teachers to "understand what their students expect from them and develop their pedagogical techniques through reflection on teaching, which will in turn enhance the complex process of teaching and learning" (Park and Lee, 2008, p. 237).

The study's second research question sought to analyze student perceptions of teaching performance in terms of lecturer national identity. To do so, the study looked at SET scores for native (Turkish) and foreign lecturers and produced similar results for both groups. As

mentioned with respect to the first research question, assuming that lecturers and students are taught to assess based on the expected criteria of the educational institution, these results suggest that identical expectations exist across cultures; being Turkish does not earn a lecturer higher ratings from Turkish students. If a cultural bias exists, it is slight, producing negligible results in the data.

As for the third research question, the findings show that there is no remarkable difference in how Turkish and foreign lecturers perceive their teaching performance. Overall, Turkish lecturers gave themselves slightly higher ratings than foreign lecturers did. The lack of a significant difference in the data may have been the result of a high awareness from both the Turkish and foreign lecturers about the school's policies and their students' expectations from them. That every educational institute has specific criteria and expectations from its teaching staff may provide a consistent basis for evaluation; teachers who are aware of these criteria could therefore implement them in the classroom, and thus they could rank themselves higher on school-issued evaluation forms. Because the present study used an assessment tool which is unique to the institution and regularly used, the lecturers could have previously familiarized themselves with the items on which they were assessed. This fact might account for the similar self-assessment scores of Turkish and foreign lecturers.

Due to several limitations, the findings of this study are neither comprehensive nor conclusive. Firstly, the study was limited by the number of its participants, which decreases the generalizability of its results. Additionally, because the self-assessment tool used in this study was developed specifically for this institution, it cannot be implemented in other schools to further corroborate the findings, which is a problem for the external validity of the study. The third limitation of this study may also lead to future areas of research: because the study relied exclusively on survey results, it could not explore the degree to which lecturers accepted or rejected the assessment ratings. Further research could utilize interview questions to triangulate the data and measure the impact that self-assessment and SET scores have on teacher development by revealing the deeper insights in teachers' heads.

## 5. REFERENCES

- D'Apollonia, S., & Abrami, P.C. (1997). Navigating student ratings of instruction. American Psychologist, 52(11), 1198-1208.*
- Arreola, R. A. (2000). Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system (2nd ed.). Bolton, MA: Anker.*
- Avery, R.J., Bryant, W.K., Kang A.M.H., & Bell, D. (2006) Electronic course evaluations: Does an online delivery system influence student evaluations? Journal of Economic Education 37, 21–37.*
- Bedard, K., & P. Kuhn (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. Economics of Education Review 27(3), 253-265.*
- Bélanger C.H. & Longden, B. (2009). The effective teacher's characteristics as perceived by students. Tertiary Education and Management 15(4), 323-340.*
- Beran, T.N., & Rokosh, J.L. (2009). The consequential validity of student ratings: What do instructors really think? Alberta Journal of Educational Research 55(4), 497-511.*
- Berk, R.A. (2005). Survey of 12 strategies to measure teaching effectiveness. International Journal of Teaching and Learning in Higher Education 17(1): 48-62.*
- Braskamp, L. A., & Ory, J. C. (1994). Assessing faculty work. San Francisco: Jossey-Bass.*
- Chen, Y., & Hoshower, L.B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. Assessment & Evaluation in Higher Education 28(1), 71-88.*



### The Relationship between Student Evaluation of Lecturer Performance and Lecturer Self-Assessment

Dommeyer, C.J., Baum, P., Hanna, R.W., & Chapman, K.S. (2004). *Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations*. *Assessment & Evaluation in Higher Education*, 29(5), 611-623

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Higher Education Quality Council of Ontario: Toronto, ON.

Griffin, B.W. (2004). *Grading leniency, grade discrepancy, and student ratings of instruction*. *Contemporary Educational Psychology* 29(4), 410-425.

Harris T. & Twiname, L. (2012). *Student evaluations of teaching: The students' perspective*. *Studies in Learning, Evaluation, Innovation and Development*, 9(1), 145-156.

Kasasen, K. & R  ty, H. (2002). *You be sure now to be honest in your assessment: Teaching and learning self-assessment*. *Social Psychology of Education* 5, 313-328.

Kelly, M. (2012) *Student evaluations of teaching effectiveness: Considerations for Ontario Universities*. Retrieved May 14, 2014 from <http://cou.on.ca/publications/academic-colleague-papers/pdfs/ac-paper---student-evaluations-may-23-12>

Martinson, D.L., & Ryan, M. (1981). *What do J-students think about teacher evaluations?* *Journalism Educator*, 36(2), 53-55.

Marsh, H.W. (1984). *Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and utility*. *Journal of Educational Psychology*, 76(5), 707-754.

Marsh, H.W. (1987). *Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research*. *International Journal of Education Research* 11(3), 253-388.

Marsh, H. W., & Roche, L. A. (2000). *Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent bystanders?* *Journal of Educational Psychology* 92(1), 202-228.

McKeachie, W.J. (1979). *Student ratings of faculty: A reprise*. *Academe* 62, 384-397.

McKeachie, W. J. (1997). *Student ratings: The validity of use*. *American Psychologist*, 52(11), 1218 - 1225.

Naftulin, D.H., J.E. Ware, & Donnelly, F.A. (1973). *The Doctor Fox Lecture: A paradigm of educational seduction*. *Journal of Medical Education* 48(7), 630-635.

Nulty, D.D. (2007). *The adequacy of response rates to online and paper surveys: What can be done?* *Assessment and Evaluation in Education*, 33(3), 301-314.

Olivares, O.J. (2001). *Student interest, grading leniency, and teacher ratings: A conceptual analysis*. *Contemporary Educational Psychology* 26(3), 382-399.

Park, G.P., & Lee, H. W. (2008). *The characteristics of effective English teachers as perceived by high school teachers and students in Korea*. *Asia Pacific Education Review*, 7(2), 236-248.

Seldin, P. (1999). *Current practices – good and bad –nationally*. In P. Seldin & Associates (Eds.), *Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions* (pp. 1–24). Bolton, MA: Anker

Spencer, K.J., & Schmelkin, L.P. (2002). *Student perspectives on teaching and its evaluations*. *Assessment and Evaluation in Higher Education*, 27(5), 397-409.

Theall, M., & Franklin, J. (2001). *Looking for bias in the all the wrong places: A search for truth or a witch hunt in student ratings of instruction?* *New Directions for Institutional Research*, 27(5), 45-56.

Van Diggelen, M., den Brok, P., & Beijgaard, D. (2012). *Teachers' use of a self-assessment procedure: The role of criteria, standards, feedback and reflection*. *Teachers and Teaching: Theory and Practice*, 19(2), 115-134.

Wheeler, A. E., & Knoop, H. R. (1982). *Self, teacher, and faculty assessments of student teacher education*. *Journal of Educational Research*, 75(3), 178-191.

